

Comparative Study on Noise-Estimation-Based Fuzzy C-Means Clustering for Image Segmentation

Cong Wang^{ID}, Member, IEEE, MengChu Zhou^{ID}, Fellow, IEEE, Witold Pedrycz^{ID}, Life Fellow, IEEE,
and ZhiWu Li^{ID}, Fellow, IEEE

Abstract—Since a noisy image has inferior characteristics, the direct use of Fuzzy C-Means (FCM) to segment it often produces poor image segmentation results. Intuitively, using its ideal value (noise-free image) benefits FCM’s robustness enhancement. Therefore, the realization of accurate noise estimation in FCM is a new and important task. To date, only two noise-estimation-based FCM algorithms have been proposed for image segmentation, that is: 1) deviation-sparse FCM (DSFCM) and 2) our earlier proposed residual-driven FCM (RFCM). In this article, we make a thorough comparative study of DSFCM and RFCM. We demonstrate that an RFCM framework can realize more accurate noise estimation than DSFCM when different types of noise are involved. It is mainly thanks to its utilization of noise distribution characteristics instead of noise sparsity used in DSFCM. We show that DSFCM is a particular case of RFCM, thus signifying that they are the same when only impulse noise is involved. With a spatial information constraint, we demonstrate RFCM’s superior effectiveness and efficiency over DSFCM in terms of supporting experiments with different levels of single, mixed, and unknown noise.

Manuscript received 23 May 2022; revised 15 August 2022; accepted 26 October 2022. This work was supported in part by the China National Postdoctoral Program for Innovative Talents under Grant BX2021249; in part by the Fellowship of China Postdoctoral Science Foundation under Grant 2021M702678; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110019; in part by the National Natural Science Foundation of China under Grant 61873342, Grant 61672400, Grant 62076189, and Grant 62206220; in part by the Recruitment Program of Global Experts; and in part by the Science and Technology Development Fund, MSAR, under Grant 0012/2019/A1. This article was recommended by Associate Editor A. F. Skarmeta Gómez. (*Corresponding authors:* MengChu Zhou; ZhiWu Li.)

Cong Wang is with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Research and Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 51800, China (e-mail: congwang0705@nwpu.edu.cn).

MengChu Zhou is with the School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China, and also with the Helen and John C. Hartmann Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: zhou@njit.edu).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, also with the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China, and also with the Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: wpedrycz@ualberta.ca).

ZhiWu Li is with the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China, and also with the Institute of Systems Engineering, Macau University of Science and Technology, Macau, China (e-mail: zhwl@xidian.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2022.3217897>.

Digital Object Identifier 10.1109/TCYB.2022.3217897

Index Terms—Comparative study, distribution characteristic, Fuzzy C-Means (FCM), image segmentation, noise estimation.

I. INTRODUCTION

SINCE its inception in 1984, a Fuzzy C-Means (FCM) algorithm has started to play a vital role in image segmentation [1], [2], [3], [4]. A common problem in its real-world application is the presence of various noise in the observed image data. The conventional FCM exhibits an inferior noise tolerance [2]. Over the past three decades, researchers have made significant efforts to enhance its robustness to noise [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Traditionally, there exist two common ways to make such improvements, that is, by incorporating spatial information of image pixels in a clustering process [5], [6], [7], [8], [9], [10] and substituting a nonrobust Euclidean distance with a flexible kernel function [11], [12], [13], [14], [15], [16], [17], [18]. Besides them, some sophisticated approaches and concepts, such as Kullback–Leibler (KL) divergence [19], [20], [21]; sparse regularization [22], [23]; morphological reconstruction [24], [25]; and gray level histograms [26], as well as preprocessing and post-processing steps, such as image pixel filtering [27]; membership filtering [28]; and label filtering [29], are often integrated into clustering environment, thus resulting in many comprehensive variants of FCMs. In brief, such existing studies make evident efforts to improve FCM’s robustness to noise by the aid of noise removal in each iteration or before and after clustering. Such noise removal shows an impact on the final attribution of a partition matrix of FCM. However, in most cases, noise removal causes the loss of effective image information. Furthermore, all the algorithms mentioned above directly use observed image data rather than their ideal values (noise-free image data) estimated from them in clustering. As a result, their segmentation results are not fully reliable and can be further improved.

Intuitively, using a noise-free image instead of its observed values (noisy image) as clustering data can greatly enhance FCM’s segmentation performance. Superior to the direct use of noise removal, integrating noise estimation into FCM is an effective way to improve FCM’s robustness, thus resulting in noise-estimation-based FCM (NFCM), where a regularization term is considered as a part of the FCM. In this sense, NFCM can not only effectively enhance FCM’s segmentation performance but also lead to FCM’s theoretical development.

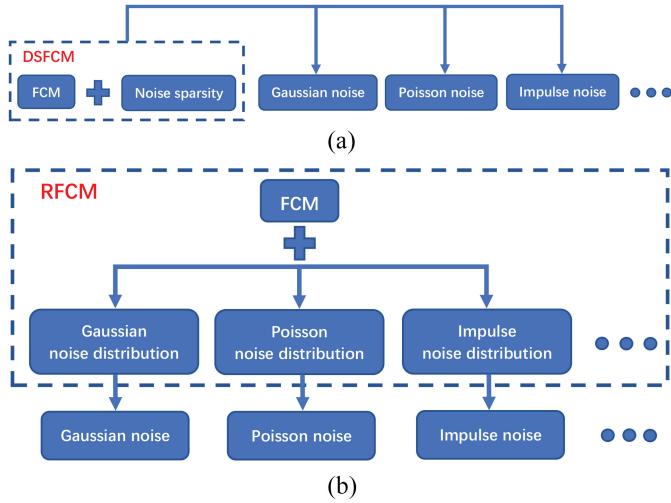


Fig. 1. Framework comparison between (a) DSFCM and (b) RFCM.

Accurate noise estimation is a very challenging task. To the best of our knowledge, there are only two attempts, namely deviation-sparse FCM (DSFCM) [30] and our residual-driven FCM (RFCM) [31], to enhance FCM's robustness effectively in virtue of a fine realization of noise estimation. The underlying contents are illustrated in Fig. 1.

To be specific, in [30], noise sparsity in an observed image is revealed under an assumption that most of image pixels have small or near zero noise/outliers. In the sequel, they use ℓ_1 -norm regularization to characterize the noise sparsity and incorporate it as a part of FCM, thus resulting in DSFCM. From the optimal solution to DSFCM, we see that a soft thresholding operation is used to realize noise estimation. In essence, DSFCM works well only when impulse noise is involved. Except impulse noise, the distribution of most types of noise does not exhibit such sparsity. It makes more sense that noise or outliers are sparse in some transformed domain. Such transforms can be Fourier transform; orthonormal wavelets [32]; translation-invariant wavelets [33]; framelets [34], [35], [36], [37]; etc. Therefore, due to the inappropriate assumption, DSFCM cannot realize accurate noise estimation and recover the ideal values from the observed image data precisely when coping with many other noise types than impulse noise.

Inspired by [30], an RFCM framework for image segmentation is proposed in our study [31]. It differs from DSFCM in the sense that we take a wide range of noise distribution characteristics used for noise estimation as a starting point and substitute multiple regularization for ℓ_1 -norm regularization, and achieves even better segmentation results. In [31], we offer an explanation of its motivation and model formulation. However, due to limited space, we do not consider a detailed theoretical derivation. Moreover, we do not reveal its superiority over DSFCM in [31].

In order to greatly improve the method in [30], this study focuses on a detailed development of RFCM and make a thorough analysis of the performance of RFCM and DSFCM. Specifically, it makes threefold contributions:

- 1) Theoretically and experimentally, we show that RFCM realizes more accurate noise estimation than DSFCM,

which stems from an exploration of noise distribution characteristics instead of noise sparsity only.

- 2) We prove that DSFCM is a particular case of RFCM. When only impulse noise is involved in an image, DSFCM is the same as RFCM.
- 3) With a spatial information constraint, we experimentally demonstrate that RFCM has a higher effectiveness and efficiency than DSFCM.

This work builds upon and substantially extends our previous results given in [31], which offers a development of RFCM, that is, a detailed derivation for RFCM's motivation and formulates the model. The interpretability of RFCM is carefully illustrated. This work expands the functionality of the method by introducing a variety of regularization terms corresponding to different types of noise. This implies that more options for the inclusion of the regularization term in the NFCM are provided. In addition, a thorough analysis of the performance of RFCM over another NFCM algorithm is covered, that is, DSFCM, which is absent in [31]. The superiority of RFCM over DSFCM is fully demonstrated for the first time. Thus, it is worth stressing that RFCM is the best NFCM algorithm so far.

For the sake of simplicity, three contributions of this study are labeled as “Contribution 1,” “Contribution 2,” and “Contribution 3,” respectively. Section III introduces the problem formulation and pertinent background. Section IV offers the proof of Contribution 1. Contribution 2 is demonstrated in Section V. As to Contribution 3, experimental results are provided in Section VI. In Section VII, some conclusions and further studies are covered.

II. RELATED WORK

Bezdek et al. [2] first proposed FCM. So far, it has evolved into the most popular fuzzy clustering algorithm. However, it cannot work well for segmenting observed (noisy) images. It has been improved by mostly considering spatial information [5], [6], [7], [8], [9], [10], kernel distances (functions) [11], [12], [13], [14], [15], [16], [17], [18], and various mathematical techniques [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. In this article, we mainly focus on the improvement of FCM with regard to its robustness to noise for image segmentation. Therefore, we introduce related work about it in this section.

A. FCM With Spatial Information

Over the past two decades, using spatial information to improve FCM's robustness achieved remarkable successes, thus resulting in many improved versions [5], [6], [7], [8], [9], [10]. For instance, Ahmed et al. [5] introduced a neighbor term into the objective function of FCM so as to improve its robustness by leaps and bounds, thus yielding FCM_S where S referred to “spatial information.” To further improve it, Chen and Zhang [6] integrated mean and median filters into a neighbor term, resulting in two FCM_S variants labeled as FCM_S1 and FCM_S2. However, their computing overhead is very high. To lower it, Szilagyi et al. [7] proposed an enhanced FCM (EnFCM) where a weighted sum image was generated

by the observed pixels and their neighborhoods. Based on it, Cai et al. [8] substituted image pixels by gray level histograms, which gives rise to fast generalized FCM (FGFCM). Although it has a high computational efficiency, more parameters are required and tuned. Krnidis and Chatzis [9] came up with a fuzzy local information C-means algorithm (FLICM) for simplifying the parameter setting in FGFCM. Nevertheless, FLICM considers only nonrobust Euclidean distance that is not applicable to arbitrary spatial information.

B. FCM With Kernel Distance

To address the serious shortcoming of FLICM [9], kernel distances (functions) are used to replace the Euclidean distance in FCM. They realize the transformation from an original data space to a new one. As a result, a collection of kernel-based FCMs has been put forward [11], [12], [13], [14], [15], [16], [17], [18]. For example, Gong et al. [11] proposed an improved version of FLICM, namely, KWFLICM, which augmented a tradeoff weighted fuzzy factor and a kernel metric into FCM. Even though it is generally robust to extensive noise, it is more time-consuming than most of existing FCMs. Zhao et al. [16] took a neighborhood weighted distance into account, thus presenting a novel algorithm called NWFCM. Although it runs faster than KWFLICM, its segmentation performance is worse. Moreover, it exhibits lower computational efficiency than other FCMs. More recently, Wang et al. [18] considered tight wavelet frames as a kernel function so as to present wavelet frame-based FCM (WFCM), which took full advantage of the feature extraction capacity of tight wavelet frames. In spite of its rarely low computational cost, its segmentation effects can be further improved by using various mathematical techniques.

C. Comprehensive FCM

To keep a sound tradeoff between performance and speed of clustering, comprehensive FCMs involving various techniques has been put forward [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. For instance, Gharieb et al. [19] presented an FCM framework based on KL divergence. It uses KL divergence to optimize the membership similarity between a pixel and its neighbors. Yet it has low clustering speed. Gu et al. [22] reported a fuzzy double C-means algorithm (FDCM) through the use of sparse representation, which addressed two datasets simultaneously, i.e., a basic feature set associated with an observed image and a feature set learned from a spare self-representation model. Overall, FDCM is robust and applicable to a wide range of image segmentation problems. However, its computational efficiency is not satisfactory. Lei et al. [28] presented a fast and robust FCM algorithm (FRFCM) by using gray level histograms and morphological gray reconstruction. In spite of its fast clustering, its performance is sometimes unstable since morphological gray reconstruction may cause the loss of useful image features. More recently, Lei et al. [26] proposed an automatic fuzzy clustering framework (AFCF) by incorporating threefold techniques, i.e., superpixel algorithms, density

peak clustering and prior entropy. It overcomes two difficulties in existing algorithms [18], [22], [28]. One is to select the number of clusters automatically. The other one is to employ superpixel algorithms and the prior entropy to improve the image segmentation performance. However, AFCF's results are unstable.

In this work, the proposed algorithm differs from all algorithms mentioned above in the sense that we take a wide range of mixed noise estimations as the starting point and directly minimize the objective function of RFCM formulated by using multiple regularization without dictionary learning [38], [39] and convolutional neural networks [40], [41], [42], [43], [44], and achieves outstanding performance in image segmentation tasks.

III. PROBLEM FORMULATION AND BACKGROUND

For a domain $\Omega = \{1, 2, \dots, K\}$, an observed image defined on Ω is denoted as $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^L : i \in \Omega\}$, where \mathbf{x}_i has L channels, that is, $\mathbf{x}_i = \{x_{il} : l = 1, 2, \dots, L\}$. To be specific, \mathbf{X} is a gray image for $L = 1$ while it is a Red–Green–Blue color image for $L = 3$. Since there exists a noise perturbation in the process of image acquisition and transmission, \mathbf{X} can be modeled as

$$\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{R} \quad (1)$$

where $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_i \in \mathbb{R}^L : i \in \Omega\}$ denotes an unknown noise-free image and $\mathbf{R} = \{\mathbf{r}_i \in \mathbb{R}^L : i \in \Omega\}$ is a noise perturbation, for example, Gaussian noise, impulse noise, and mixed Gaussian and impulse noise. Formally speaking, $\tilde{\mathbf{X}}$ is an ideal value of \mathbf{X} and \mathbf{R} is viewed as a residual (difference) between \mathbf{X} and $\tilde{\mathbf{X}}$.

A. Fuzzy C-Means

FCM is often used to divide an observed image into several nonoverlapping clusters. Given is an observed image \mathbf{X} , FCM is used to segment it by minimizing the objective function

$$J^{\text{FCM}}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (2)$$

where $\mathbf{U} = [u_{ij}]_{K \times c}$ is a partition matrix with constraints $\sum_{j=1}^c u_{ij} = 1$ for $\forall i$ and $0 < \sum_{i=1}^K u_{ij} < K$ for $\forall j$, $\mathbf{V} = \{\mathbf{v}_j \in \mathbb{R}^L : j \in \{1, 2, \dots, c\}\}$ denotes a prototype set, $\|\cdot\|$ represents a Euclidean distance, and m is a fuzzification exponent. m assumes values greater than 1.

B. Noise-Estimation-Based FCM

Due to the presence of a noise perturbation, the direct use of an observed image gives rise to bad segmentation results. From (1), we see that its noise-free image can be computed when the noise is accurately estimated. Intuitively, using its noise-free image instead of it as data to be clustered benefits FCM's robustness improvement. Therefore, an ideal variant of FCM's objective function (2) for image segmentation is expressed in the form

$$J^{\text{FCM}}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\tilde{\mathbf{x}}_i - \mathbf{v}_j\|^2. \quad (3)$$

By (1), we can transform (3) into

$$J^{\text{FCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2. \quad (4)$$

To make (4) executed, the main task is to realize accurate estimation of \mathbf{R} . Therefore, it is reasonable to impose a regularization term $\Gamma(\mathbf{R})$ to (4), which characterizes some properties of \mathbf{R} . As a result, NFCM is proposed. The augmented objective function is defined as

$$\begin{aligned} J^{\text{NFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) &= \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2 + \boldsymbol{\beta} \cdot \Gamma(\mathbf{R}) \\ &= \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2 + \sum_{l=1}^L \beta_l \Gamma(\mathbf{R}_l). \end{aligned} \quad (5)$$

In (5), \mathbf{R} is rewritten as $\{\mathbf{R}_l \in \mathbb{R}^K : l = 1, 2, \dots, L\}$ with $\mathbf{R}_l = (r_{1l}, r_{2l}, \dots, r_{Kl})^T$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_l, \dots, \beta_L)$ is a positive parameter vector that is used to strike a balance between $\Gamma(\mathbf{R}) = (\Gamma(\mathbf{R}_1), \dots, \Gamma(\mathbf{R}_l), \dots, \Gamma(\mathbf{R}_L))^T$ and J^{FCM} in (4). The regularization term $\Gamma(\mathbf{R})$ guarantees that the solution accords with the degradation process of minimizing (5). As (5) indicates, the main task involved in NFCM is to determine the expression of $\Gamma(\mathbf{R}_l)$ based on an analysis of the property of the l th residual channel \mathbf{R}_l .

1) *Deviation-Sparse FCM*: In [30], an achievable NFCM version is proposed for image segmentation, which is known as DSFCM. It is modeled based on an assumption that most of the pixels in an observed image \mathbf{X} have small or near zero noise/outliers. As a result, the residual \mathbf{R} shows strong sparsity. This study focuses on characterizing noise sparsity by using an ℓ_1 -norm regularization term on residual

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1} = \sum_{i=1}^K |r_{il}| \quad (6)$$

where $\|\cdot\|_{\ell_1}$ denotes an ℓ_1 vector norm and $|\cdot|$ denotes an absolute value operation. By substituting (6) into (5), DSFCM's objective function is expressed as

$$J^{\text{DSFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2 + \sum_{l=1}^L \beta_l \sum_{i=1}^K |r_{il}|. \quad (7)$$

For image segmentation, an arbitrary image pixel is close to its neighbors. Therefore, neighbor or spatial information is usually introduced into FCM, which usually makes a critical impact on the final attribution of a partition matrix \mathbf{U} . By combining neighbor information and DSFCM, an enhanced version, labeled as DSFCM_N, is given

$$\begin{aligned} J^{\text{DSFCM_N}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) &= \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \left(\sum_{n \in \mathcal{N}_i} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_j\|^2}{1 + d_{ni}} \right) \\ &\quad + \sum_{l=1}^L \beta_l \sum_{i=1}^K \sum_{n \in \mathcal{N}_i} \frac{|r_{nl}|}{1 + d_{ni}}. \end{aligned} \quad (8)$$

In (8), an image pixel is loosely represented by its corresponding index even though this is not ambiguous. n is a neighbor of i , \mathcal{N}_i represents a local window centralized in i including i , and d_{ni} denotes a Euclidean distance between n and i . The minimization of (8) is realized in [30, Algorithm 1].

2) *Residual-Driven FCM*: In [31], a flexible regularization term on residual is introduced into NFCM, which is modeled based on specified noise distribution characteristics rather than noise sparsity. In other words, in the presence of different types of noise, we can model a regularization term corresponding to each noise type. Some examples for common noise types are summarized in Table I in the next section. As a result, an RFCM framework is proposed. Although RFCM's expression is the same as (5), it offers more options for a regularization term in NFCM. For example, ℓ_2 -norm regularization [45], [46] is used for Gaussian noise, ℓ_1 -norm regularization is applicable to impulse noise [47], [48], and Csiszár's I-divergence [49], [50] can be regarded as a regularization term for Poisson noise.

In real-world applications, an observed image often suffers from mixed or unknown noise. Consequently, regularization terms for single noise are usually not the best option. In [31], by an analysis of a wide range of mixed noise distributions, a weighted ℓ_2 -norm regularization term is proposed

$$\Gamma(\mathbf{R}_l) = \|\mathbf{W}_l \circ \mathbf{R}_l\|_{\ell_2}^2 = \sum_{i=1}^K |w_{il} r_{il}|^2 \quad (9)$$

where $\|\cdot\|_{\ell_2}$ denotes an ℓ_2 vector norm, \circ stands for an elementwise product, and $\mathbf{W}_l = (w_{1l}, w_{2l}, \dots, w_{Kl})^T$ is a weight vector generated by an exponential function

$$w_{il} = \exp(-\zeta \cdot r_{il}^2) \quad (10)$$

with ζ being a parameter that controls the decreasing rate of w_{il} . We briefly clarify the effect of weighting, that is, transforming a mixed noise distribution into a Gaussian-like one, which makes ℓ_2 -norm regularization easily used. Small residuals basically obey a Gaussian distribution and they can remain unchanged. That is to say, they should be associated with weights whose values are close to 1. The remaining residuals should be assigned with smaller weights to reduce the heavy tail of the mixed noise distribution. More details on the effect of weighting can be found in [31].

By substituting (9) into (5), a universal NFCM model, that is, RFCM with weighted ℓ_2 -norm regularization, namely, WRFCM, is proposed for coping with mixed or even unknown noise

$$\begin{aligned} J^{\text{WRFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{W}) &= \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2 \\ &\quad + \sum_{l=1}^L \beta_l \sum_{i=1}^K |w_{il} r_{il}|^2. \end{aligned} \quad (11)$$

Motivated by [30], spatial information is considered in WRFCM. As a result, an augmented version of (11) is expressed as

$$J^{\text{WRFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{W}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \left(\sum_{n \in \mathcal{N}_i} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_j\|^2}{1 + d_{ni}} \right) + \sum_{l=1}^L \beta_l \sum_{i=1}^K \sum_{n \in \mathcal{N}_i} \frac{|w_{nl} r_{nl}|^2}{1 + d_{ni}}. \quad (12)$$

The minimization of (12) is realized in [31, Algorithm 2].

IV. PROOF OF CONTRIBUTION 1

By (1), a noise-free image $\tilde{\mathbf{X}}$ and the residual \mathbf{R} can be converted to each other. Motivated by a variational model for image restoration [46], a regularization term $\Gamma(\mathbf{R}_l)$ on \mathbf{R}_l can be equivalently converted to a fidelity term $\Psi(\tilde{\mathbf{X}}_l)$ on $\tilde{\mathbf{X}}_l$, that is

$$\Gamma(\mathbf{R}_l) \Leftrightarrow \Psi(\tilde{\mathbf{X}}_l) \text{ for } l = 1, 2, \dots, L. \quad (13)$$

The fidelity term $\Psi(\tilde{\mathbf{X}}_l)$ usually penalizes the discrepancy between $\tilde{\mathbf{X}}_l$ and \mathbf{X}_l . Its formulation leans on an assumed noise distribution. Since the residual \mathbf{R}_l stands for such a discrepancy, it is easy to understand that the regularization term $\Gamma(\mathbf{R}_l)$ is associated with the assumed noise distribution. From (13), we see that once $\Psi(\tilde{\mathbf{X}}_l)$ is given, the corresponding regularization model $\Gamma(\mathbf{R}_l)$ is determined. Therefore, before formulating $\Gamma(\mathbf{R}_l)$, a key step is to derive $\Psi(\tilde{\mathbf{X}}_l)$ by using a maximum a posteriori (MAP) method and a likelihood function. The details are formulated as follows.

The MAP estimate of $\tilde{\mathbf{X}}_l$ can be found by maximizing the conditional posterior probability $\mathbf{P}(\tilde{\mathbf{X}}_l | \mathbf{X}_l)$, that is, the probability that $\tilde{\mathbf{X}}_l$ occurs when \mathbf{X}_l is observed. Based on Bayes' theorem, we have

$$\mathbf{P}(\tilde{\mathbf{X}}_l | \mathbf{X}_l) = \frac{\mathbf{P}(\tilde{\mathbf{X}}_l) \cdot \mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l)}{\mathbf{P}(\mathbf{X}_l)}. \quad (14)$$

By taking the negative logarithm of (14) and ignoring $\mathbf{P}(\tilde{\mathbf{X}}_l)$ unassociated with \mathbf{R} and a constant $\mathbf{P}(\mathbf{X}_l)$, the estimate is acquired by solving the following optimization problem:

$$\max_{\tilde{\mathbf{X}}_l} \log \mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l) \Leftrightarrow \min_{\tilde{\mathbf{X}}_l} -\log \mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l). \quad (15)$$

The likelihood $\mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l)$ is used to yield a fidelity term on $\tilde{\mathbf{X}}_l$, which measures the discrepancy between $\tilde{\mathbf{X}}_l$ and \mathbf{X}_l . Its choice depends on specific noise characteristics.

To make the determination process of $\mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l)$ easily understood, we take the presence of additive white Gaussian noise as a case study. In the sequel, we derive the expression form of $\Psi(\tilde{\mathbf{X}}_l)$ and $\Gamma(\mathbf{R}_l)$. Then, RFCM for Gaussian noise is presented. The derivation procedure is detailed as follows.

Consider an observed image \mathbf{X} and the presence of additive white Gaussian noise. For $l = 1, 2, \dots, L$, \mathbf{R}_l obeys an independent normal distribution \mathbf{N} of zero mean and variance σ_l^2 , that is

$$\mathbf{R}_l \sim \mathbf{N}(0, \sigma_l^2). \quad (16)$$

Thus, we arrive at the likelihood

$$\begin{aligned} \mathbf{P}(\mathbf{X}_l | \tilde{\mathbf{X}}_l) &= \mathbf{P}(\mathbf{R}_l | \tilde{\mathbf{X}}_l) \\ &\simeq \exp\left(-\frac{1}{2\sigma_l^2} (\mathbf{X}_l - \tilde{\mathbf{X}}_l)^T (\mathbf{X}_l - \tilde{\mathbf{X}}_l)\right). \end{aligned} \quad (17)$$

Using maximum likelihood, we take the negative logarithm of the normal distribution (17), that is

$$\begin{aligned} -\log \mathbf{P}(\mathbf{R}_l | \tilde{\mathbf{X}}_l) &\propto \frac{1}{2\sigma_l^2} (\mathbf{X}_l - \tilde{\mathbf{X}}_l)^T (\mathbf{X}_l - \tilde{\mathbf{X}}_l) \\ &\propto \frac{1}{2\sigma_l^2} \|\mathbf{X}_l - \tilde{\mathbf{X}}_l\|_{\ell_2}^2. \end{aligned} \quad (18)$$

By ignoring the coefficient in (18), we obtain a fidelity term on $\tilde{\mathbf{X}}_l$

$$\Psi(\tilde{\mathbf{X}}_l) = \|\mathbf{X}_l - \tilde{\mathbf{X}}_l\|_{\ell_2}^2 = \sum_{i=1}^K |x_{il} - \tilde{x}_{il}|^2. \quad (19)$$

By (1), (13), and (19), the regularization term on \mathbf{R}_l is expressed as

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_2}^2 = \sum_{i=1}^K |r_{il}|^2. \quad (20)$$

By substituting (20) into (5), RFCM's objective function associated with Gaussian noise is formulated

$$J^{\text{RFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|\mathbf{x}_i - \mathbf{r}_i - \mathbf{v}_j\|^2 + \sum_{l=1}^L \beta_l \sum_{i=1}^K |r_{il}|^2 \quad (21)$$

where β_l is associated with σ_l .

Remark 1: By (16), we see that the variance of Gaussian noise \mathbf{R}_l is σ_l^2 , that is

$$\frac{1}{K} \sum_{i=1}^K r_{il}^2 = \frac{1}{K} \|\mathbf{R}_l\|_{\ell_2}^2 = \sigma_l^2. \quad (22)$$

The construction of $\Gamma(\mathbf{R}_l)$ in (20) aims to characterize Gaussian noise distribution (22). By observing (20) and (22), we find that they are consistent in theory, that is, producing an ℓ_2 -norm term on residual. As a result, ℓ_2 -norm regularization is more suitable to characterize a Gaussian noise distribution than ℓ_1 -norm regularization used in DSFCM.

However, such a remark is not rigorous. To convincingly demonstrate that RFCM can realize more accurate noise estimation than DSFCM, we need to determine that the residual results produced by RFCM are closer to true Gaussian noise than DSFCM's. Formally speaking, the objective functions of DSFCM and RFCM have been previously presented as (7) and (21). By applying the Lagrangian multiplier method to minimize (7) and (21), we can acquire the expressions of a partition matrix \mathbf{U} , prototypes \mathbf{V} and residuals \mathbf{R} . RFCM and DSFCM have the same expressions of \mathbf{U} and \mathbf{V} , which can be found as (15) and (16) in [30]. We here focus on their expressions of \mathbf{R} , which are respectively presented as

$$r_{il}^{\text{RFCM}} = \frac{\sum_{j=1}^c u_{ij}^m (x_{il} - v_{jl})}{\sum_{j=1}^c u_{ij}^m + \beta_l} \quad (23)$$

$$r_{il}^{\text{DSFCM}} = \frac{\mathcal{S}\left(\sum_{j=1}^c u_{ij}^m (x_{il} - v_{jl}), \beta_l/2\right)}{\sum_{j=1}^c u_{ij}^m} \quad (24)$$

where \mathcal{S} is a soft-thresholding operator [51], [52], [53] defined as

$$\mathcal{S}(\theta, \delta) = \text{sign}(\theta) \cdot \max\{|\theta| - \delta, 0\}.$$

For convenience, we denote true Gaussian noise located at the i th pixel in the l th channel as $r_{il}^* = x_{il} - \tilde{x}_{il}$. For the l th pixel channel, we measure the discrepancies between true Gaussian noise and residual results of RFCM and DSFCM, respectively. We denote them as $\vartheta_l^{\text{RFCM}} = \{\vartheta_{il}^{\text{RFCM}} : i \in \Omega\}$ and $\vartheta_l^{\text{DSFCM}} = \{\vartheta_{il}^{\text{DSFCM}} : i \in \Omega\}$, respectively

$$\vartheta_{il}^{\text{RFCM}} = |r_{il}^* - r_{il}^{\text{RFCM}}|, \quad \vartheta_{il}^{\text{DSFCM}} = |r_{il}^* - r_{il}^{\text{DSFCM}}|.$$

If $\vartheta_{il}^{\text{DSFCM}} > \vartheta_{il}^{\text{RFCM}}$, r_{il}^{RFCM} is closer to r_{il}^* than r_{il}^{DSFCM} . We denote the discrepancy between $\vartheta_l^{\text{DSFCM}}$ and $\vartheta_l^{\text{RFCM}}$ as $\Theta_l = \{\Theta_{il} : i \in \Omega\}$, which is defined as

$$\Theta_{il} = \vartheta_{il}^{\text{DSFCM}} - \vartheta_{il}^{\text{RFCM}}.$$

If the expectation $\mathbf{E}(\Theta_l) > 0$, the average residual results of RFCM are closer to true Gaussian noise than DSFCM's. Then, we can confirm that RFCM realizes more accurate noise estimations than DSFCM.

Since an alternating iteration scheme [2] is used to update u_{ij} , v_{jl} and r_{il} , for all cases $c \geq 2$, it is hard to illustrate $\mathbf{E}(\Theta_l) > 0$. For the sake of a simple comparison, let us consider the case in terms of two clusters, that is, $c = 2$. It is since, in this case, two membership degrees of an arbitrary pixel belonging to two clusters are significantly different, that is, extremely approximate to 0 or 1. To do so, we provide Theorem 1 and its proof.

Theorem 1: Given an observed image X , assume that it is corrupted with Gaussian noise \mathbf{R} with $\mathbf{R}_l \sim \mathbf{N}(0, \sigma_l^2)$. For $c = 2$, we have that the expectation $\mathbf{E}(\Theta_l) > 0$ if and only if $\beta_l > 0.76\sigma_l$. Then, RFCM realizes more accurate Gaussian noise estimation than DSFCM.

Proof: For $c = 2$, there usually exist two deviations between v_{jl} and \tilde{x}_{il} , which are denoted as Δ_{1l} and Δ_{2l} , respectively. Thus, we have

$$v_{1l} = \tilde{x}_{il} - \Delta_{1l}, \quad v_{2l} = \tilde{x}_{il} - \Delta_{2l}.$$

Assume that x_{il} belongs to the first cluster. The known conditions contain: $u_{i1} \approx 1$, $u_{i2} \approx 0$, Δ_{1l} is small positive number or near 0 while Δ_{2l} is much larger than Δ_{1l} , and $\beta_l \gg \Delta_{1l}$. Thus, we have

$$\begin{aligned} r_{il}^{\text{RFCM}} &\approx \frac{x_{il} - v_{1l}}{1 + \beta_l} = \frac{r_{il}^* + \Delta_{1l}}{1 + \beta_l} \\ r_{il}^{\text{DSFCM}} &\approx \text{sign}(r_{il}^* + \Delta_{1l}) \cdot \max\{|r_{il}^* + \Delta_{1l}| - \beta_l/2, 0\}. \end{aligned}$$

Since a Gaussian distribution is symmetric about $r_{il}^* = 0$, the proof is similar in both cases, that is, $r_{il}^* < 0$ and $r_{il}^* \geq 0$. Thus, we only focus on $r_{il}^* \geq 0$. Then, we have

$$\vartheta_{il}^{\text{RFCM}} = \frac{\beta_l r_{il}^*}{1 + \beta_l} - \frac{\Delta_{1l}}{1 + \beta_l} \quad (25)$$

$$\vartheta_{il}^{\text{DSFCM}} = \begin{cases} r_{il}^*, & r_{il}^* < \frac{\beta_l}{2} - \Delta_{1l} \\ \frac{\beta_l}{2} - \Delta_{1l}, & r_{il}^* \geq \frac{\beta_l}{2} - \Delta_{1l} \end{cases} \quad (26)$$

For $r_{il}^* < ([\beta_l]/2) - \Delta_{1l}$, we have

$$\Theta_{il} = r_{il}^* - \frac{\beta_l r_{il}^*}{1 + \beta_l} + \frac{\Delta_{1l}}{1 + \beta_l}. \quad (27)$$

For $r_{il}^* \geq ([\beta_l]/2) - \Delta_{1l}$, we have

$$\Theta_{il} = \frac{\beta_l}{2} - \Delta_{1l} - \frac{\beta_l r_{il}^*}{1 + \beta_l} + \frac{\Delta_{1l}}{1 + \beta_l}. \quad (28)$$

As mentioned above, Δ_{1l} is small or near 0, and $\beta_l \gg \Delta_{1l}$. Let us consider two extreme conditions, that is, $\Delta_{1l} \rightarrow 0$ and $([\beta_l]/[1 + \beta_l]) \rightarrow 1$. Thus, for $r_{il}^* < ([\beta_l]/2)$, (27) is rewritten as

$$\Theta_{il} = 0. \quad (29)$$

For $r_{il}^* \geq ([\beta_l]/2)$, (28) is rewritten as

$$\Theta_{il} = \frac{\beta_l}{2} - r_{il}^*. \quad (30)$$

Given a random variable $\xi(x)$, its expectation is defined as

$$\mathbf{E}(\xi) = \int_{-\infty}^{+\infty} x \mathbf{P}(x) dx. \quad (31)$$

By (31), we can acquire the expectation of Θ_l . For $r_{il}^* < ([\beta_l]/2)$, we have

$$\mathbf{E}(\Theta_l) = 0. \quad (32)$$

In this case, both algorithms have the same or similar noise estimation. For $r_{il}^* \geq ([\beta_l]/2)$, we have

$$\begin{aligned} \mathbf{E}(\Theta_l) &= \frac{\beta_l}{2} - \int_{\frac{\beta_l}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(r_{il}^*)^2}{2\sigma_l^2}\right) r_{il}^* dr_{il}^* \\ &= \frac{\beta_l}{2} - \frac{\sigma_l}{\sqrt{2\pi}} \exp\left(-\frac{\beta_l^2}{8\sigma_l^2}\right). \end{aligned} \quad (33)$$

Since β_l is associated with σ_l , we can define it as

$$\beta_l = 2\tau\sigma_l \quad (34)$$

where τ is a positive parameter. Thus, (33) is equivalently converted into

$$\begin{aligned} \mathbf{E}(\Theta_l) &= \tau\sigma_l - \frac{\sigma_l}{\sqrt{2\pi}} \exp\left(-\frac{\tau^2}{2}\right) \\ &= \sigma_l \left(\tau - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau^2}{2}\right) \right). \end{aligned} \quad (35)$$

In order for $\mathbf{E}(\Theta_l)$ in (35) to be greater than 0, we only require $f(\tau) = \tau - (1/\sqrt{2\pi}) \exp(-[\tau^2]/2) > 0$. By using a bisection method [54], we have that $f(\tau) > 0$ if and only if $\tau > 0.38$. According to (34), we have $\beta_l > 0.76\sigma_l$. That is to say, if and only if $\beta_l > 0.76\sigma_l$, RFCM can realize more accurate noise estimation than DSFCM. ■

Remark 2: By Theorem 1, we see that $\beta_l > 0.76\sigma_l$ is a prerequisite for RFCM to be superior to DSFCM in theory. Let us illustrate its underlying essence. Assume that true Gaussian noise is covered in a set \mathbb{S} . $|\mathbb{S}|$ denotes the cardinality of \mathbb{S} , which represents the amount of true Gaussian noise. Here, we have $|\mathbb{S}| = K$. According to the probability density function of a Gaussian distribution, it is clear that most of noise is small or near zero while a small proportion of noise is large.

TABLE I
DATA REGULARIZATION MODELS

Data Regularization Functions	Noise Types
$\Gamma(\mathbf{R}_l) = \ \mathbf{R}_l\ _{\ell_2}^2$	Gaussian noise [45], [46]
$\Gamma(\mathbf{R}_l) = \ \mathbf{R}_l\ _{\ell_0}$	impulse noise [47], [48]
$\Gamma(\mathbf{R}_l) = \ \mathbf{R}_l\ _{\ell_\infty}$	uniform noise [56], [57]
$\Gamma(\mathbf{R}_l) = \langle -\mathbf{R}_l \circ \log(\mathbf{X}_l - \mathbf{R}_l), \mathbf{1} \rangle$	Poisson noise [49], [50]
$\Gamma(\mathbf{R}_l) = \langle \log(\mathbf{X}_l - \mathbf{R}_l) + \mathbf{X}_l \circ \frac{1}{\mathbf{X}_l - \mathbf{R}_l}, \mathbf{1} \rangle$	Gamma noise [58], [59]
$\Gamma(\mathbf{R}_l) = \langle \log(\mathbf{X}_l - \mathbf{R}_l) + \mathbf{X}_l \circ \mathbf{X}_l \circ \frac{1}{2\mathbf{X}_l - 2\mathbf{R}_l}, \mathbf{1} \rangle$	Rayleigh noise [60], [61]
$\Gamma(\mathbf{R}_l) = \ \mathbf{Z}_l - \mathbf{R}_l\ _{\ell_2}^2, s.t. \ \mathbf{Z}_l\ _{\ell_0} \leq \kappa$	Mixed Gaussian and impulse noise [62]
$\Gamma(\mathbf{R}_l) = \ \mathbf{W}_l \circ \mathbf{R}_l\ _{\ell_2}^2$	Mixed or unknown noise (9)

Therefore, we can divide Gaussian noise into two subsets, that is, $\mathbb{S}_1 = \{r_{il}^* : |r_{il}^*| < ([\beta_l]/2)\}$ and $\mathbb{S}_2 = \{r_{il}^* : |r_{il}^*| \geq ([\beta_l]/2)\}$. We assume that \mathbb{S}_1 collects most of Gaussian noise being small or near zero. The remaining noise is put into \mathbb{S}_2 . Hence, we have the following relationship:

$$\frac{|\mathbb{S}_1|}{|\mathbb{S}|} \in (0.5, 1). \quad (36)$$

Consider that $\beta_l = 0.76\sigma_l$ and $\mathbb{S}_1 = \{r_{il}^* : |r_{il}^*| < 0.38\sigma_l\}$. By referring to the numerical table of a normal distribution function [55], we obtain

$$\Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.38\sigma_l} \exp\left(-\frac{(r_{il}^*)^2}{2}\right) dr_{il}^* = 0.648.$$

In the sequel, we obtain

$$\begin{aligned} \frac{|\mathbb{S}_1|}{|\mathbb{S}|} &= \frac{1}{\sqrt{2\pi}} \int_{-0.38\sigma_l}^{0.38\sigma_l} \exp\left(-\frac{(r_{il}^*)^2}{2}\right) dr_{il}^* \\ &= 0.648 - (1 - \Phi) \\ &= 0.296. \end{aligned}$$

Thus, when $\beta_l > 0.76\sigma_l$, we have $(|\mathbb{S}_1|/|\mathbb{S}|) \in (0.296, 1)$. Obviously, (36) is always satisfied. Therefore, when Gaussian noise is involved in real-world applications, RFCM can always realize more accurate noise estimation than DSFCM.

By following the above case, we derive the regularization terms associated with typical noise types, as summarized in Table I. Obviously, RFCM realizes accurate noise estimations in the presence of different types of noise. It has more expressive forms and is more selective than DSFCM. Therefore, it is concluded that RFCM is more applicable to different types of noise than DSFCM.

Next, we experimentally show that RFCM realizes more accurate noise estimation than DSFCM. We use an example, as portrayed in Fig. 2, to elaborate this.

As shown in Fig. 2, in presence of Gaussian and mixed noise, the noise estimation of DSFCM in Fig. 2(c) and (i) is not accurate and even far from the true one in Fig. 2(b) and (h). In contrast, DSFCM works well for impulse noise estimation, as shown in Fig. 2(f), which is since a low level of impulse noise exhibits sparsity. Superior to it, RFCM achieves better noise estimation results, as shown in Fig. 2(d), (g), and (j). Moreover, it achieves better results on noise-suppression and feature-preserving than DSFCM does.

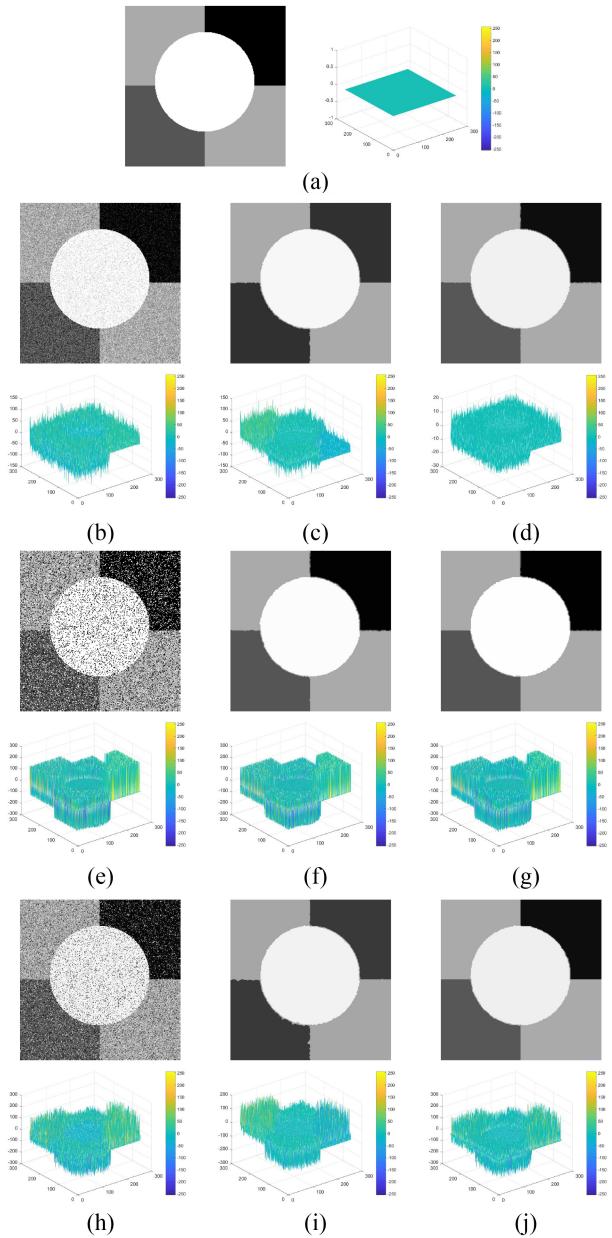


Fig. 2. Noise estimation comparison between DSFCM and RFCM. Row 1 shows an noise-free image. In other rows, from left to right: observed images [from top to bottom: Gaussian noise with level 30, impulse noise with 30%, and a mixture of Poisson, Gaussian (level 30), and impulse noise (level 20%)] and the results of DSFCM and RFCM.

V. PROOF OF CONTRIBUTION 2

Impulse noise is a common type of noise in image acquisition and transmission, which mainly results from faulty sensors or analog-to-digital converter errors. Since scratches in photographs and video sequences are sparse, they are sometimes regarded as a type of impulse noise. However, they are difficult to be removed, since their intensities are indistinguishable from those of their neighbors and the distribution of damaged image pixels is random. Generally speaking, two types of impulse noise are often encountered, that is: 1) salt-and-pepper and 2) random-valued impulse noise [47], [48]. For the l th pixel channel, we denote the minimum and maximum

pixel values as \tilde{x} and \hat{x} , respectively. Moreover, the clean and observed pixel values at location i are represented as \tilde{x}_{il} and x_{il} , respectively.

Salt-and-Pepper Impulse Noise: A certain percentage of pixels is altered to be either \tilde{x} or \hat{x}

$$x_{il} = \begin{cases} \tilde{x}, & \text{with probability } \zeta/2 \\ \hat{x}, & \text{with probability } \zeta/2 \\ \tilde{x}_{il}, & \text{with probability } 1 - \zeta. \end{cases} \quad (37)$$

Random-Valued Impulse Noise: A certain percentage of pixels are altered to take on a uniform random number $x'_{il} \in [\tilde{x}, \hat{x}]$

$$x_{il} = \begin{cases} x'_{il}, & \text{with probability } \zeta \\ \tilde{x}_{il}, & \text{with probability } 1 - \zeta \end{cases} \quad (38)$$

where ζ in (37) and (38) stands for an impulse noise level, which implies that a percentage ζ of pixels in an observed image are corrupted.

Theorem 2: Given an observed image X , assume that it is polluted with impulse noise only, that is, satisfying (37) or (38). Then, when segmenting X in the l th pixel channel, RFCM has the same regularization term as DSFCM, which is formulated as

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1}.$$

In other words, DSFCM is a special case of RFCM if only impulse noise is embedded in an image.

Proof: By solving (15), we can acquire the MAP estimate of \tilde{X}_l . From the definition of impulse noise in (37) and (38), we have

$$\mathbf{P}(X_l | \tilde{X}_l) = 1 - \zeta = 1 - \frac{\|X_l - \tilde{X}_l\|_{\ell_0}}{K} \quad (39)$$

where $\|\cdot\|_{\ell_0}$ denotes an ℓ_0 vector norm, which counts the number of nonzero entries in a vector. By (39), we see that the likelihood $\mathbf{P}(X_l | \tilde{X}_l)$ is negatively correlated with $\|X_l - \tilde{X}_l\|_{\ell_0}$, that is

$$\mathbf{P}(X_l | \tilde{X}_l) \propto -\frac{1}{K} \|X_l - \tilde{X}_l\|_{\ell_0}.$$

Since the likelihood $\mathbf{P}(X_l | \tilde{X}_l)$ is not directly related to the value of variable \tilde{X}_l , we have

$$-\log \mathbf{P}(X_l | \tilde{X}_l) \propto \frac{1}{K} \|X_l - \tilde{X}_l\|_{\ell_0}. \quad (40)$$

By ignoring the coefficient in (40), we obtain a fidelity term on \tilde{X}_l

$$\Psi(\tilde{X}_l) = \|X_l - \tilde{X}_l\|_{\ell_0}. \quad (41)$$

By (1) and (13), the regularization term on \mathbf{R}_l corresponding to (41) is expressed as

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1}. \quad (42)$$

In essence, $\Gamma(\mathbf{R}_l)$ in (42) characterizes the sparsity of impulse noise. By substituting (42) into (5), RFCM's objective function associated with impulse noise is expressed as

$$J^{\text{RFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|x_i - r_i - v_j\|^2 + \sum_{l=1}^L \beta_l \sum_{i=1}^K |r_{il}|_0 \quad (43)$$

with

$$|r_{il}|_0 = \begin{cases} 1, & r_{il} \neq 0 \\ 0, & r_{il} = 0. \end{cases}$$

In an implementation, due to the loss of amplitude characteristics of the ℓ_0 norm, it is hard to minimize (42) directly. Moreover, since the ℓ_1 norm is a convex envelope of the ℓ_0 norm, both of them have good approximate characteristics [63]. Therefore, the ℓ_1 norm approximately replacing ℓ_0 norm is often used to characterize the sparsity of impulse noise

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1}. \quad (44)$$

By substituting (44) into (5), the RFCM's objective function associated with impulse noise is rewritten as

$$J^{\text{RFCM}}(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^K \sum_{j=1}^c u_{ij}^m \|x_i - r_i - v_j\|^2 + \sum_{l=1}^L \beta_l \sum_{i=1}^K |r_{il}|. \quad (45)$$

Obviously, (45) is the same as DSFCM's objective function (18). Therefore, DSFCM is a particular case of an RFCM framework. ■

Remark 3: If only impulse noise is involved in an image, DSFCM is the same as RFCM. In addition, by (42) and (44), we see that there are two choices of regularization terms suitable for impulse noise estimation, that is, $\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_0}$ and $\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1}$.

VI. COMPARATIVE EXPERIMENTS

To demonstrate that RFCM outperforms DSFCM, we make visual and quantitative comparisons between them. With spatial information constraint, the augmented version of DSFCM, labeled as DSFCM_N, that is, (8), is used. We choose WRFCM, that is, (12) as a representative of RFCM, since it can realize accurate estimation in presence of fairly complicated mixed or unknown noise. The implementation of such two algorithms is illustrated in [30, Algorithm 1] and [31, Algorithm 2]. Their parameter settings can be found in [30] and [31]. We carry out extensive experiments involving synthetic and real-world images with different levels of single, mixed, and unknown noise. To evaluate the segmentation results of the two algorithms, we adopt three objective indicators, that is: 1) segmentation accuracy (SA) [11]; 2) Sorenson–Dice similarity (SDS) [64]; and 3) Matthews correlation coefficient (MCC) [65], [66]. The best results are displayed in a bold face.

A. Comparison Results on Synthetic Images

In the first experiment, five synthetic images are tested, as shown in Fig. 3. We impose different types of noise with a variety of levels on them, including Gaussian noise (abbreviated as “Gau.” with standard deviation), impulse noise (abbreviated as ‘imp.’ with probability) and a mixture of Poisson, Gaussian, and impulse noise (abbreviated as “mix.”). In particular, as a typical case of impulse noise, salt and pepper impulse noise is used in this experiment. The average



Fig. 3. Five synthetic images.

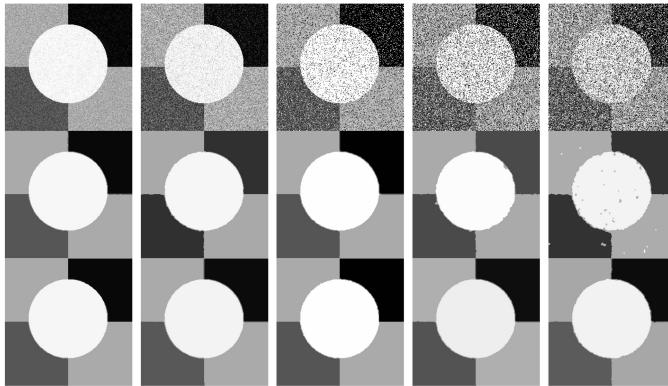


Fig. 4. Segmentation results on a synthetic image. For top to bottom: observed images and results of DSFCM_N and WRFCM. From left to right: Gaussian noise of level 20, Gaussian noise of level 40, impulse noise of level 20%, impulse noise of level 40%, and mixture of level (40, 40%).

TABLE II
AVERAGE COMPARISON RESULTS (%) ON SYNTHETIC IMAGES

Noise	DSFCM_N			WRFCM		
	SA	SDS	MCC	SA	SDS	MCC
Gau. (20)	99.311	99.786	99.386	99.405	99.787	99.389
Gau. (30)	98.982	99.740	99.256	99.304	99.785	99.386
Gau. (40)	98.932	99.736	99.242	99.252	99.770	99.342
imp. (20%)	99.368	99.781	99.579	99.382	99.768	99.335
imp. (30%)	99.042	99.784	99.379	99.085	99.712	99.185
imp. (40%)	98.539	99.537	98.651	98.866	99.711	99.171
mix. (20,20%)	99.243	99.729	99.225	99.323	99.763	99.322
mix. (30,30%)	98.973	99.651	98.998	99.255	99.761	99.114
mix. (40,40%)	98.123	99.302	97.995	99.197	99.639	98.966

comparison results of DSFCM_N and WRFCM are presented in Table II. Some representative examples are portrayed in Fig. 4.

As Table II indicates, WRFCM is basically superior to DSFCM_N in all evaluation indicators. WRFCM's average SA improvement over DSFCM is 0.29% when segmenting five synthetic images. To be specific, both DSFCM_N and WRFCM exhibit satisfactory segmentation results in presence of low levels of noise. However, with the increase of the noise level, DSFCM_N becomes incompetent for addressing image segmentation tasks. The visual inspection of Fig. 4 also confirms this finding. The presence of these results is due to the following reasons: when noise intensity and density are low, noise is almost sparse, ℓ_1 -norm regularization can work well since it is applicable to characterize the noise sparsity; however, when they become high, noise sparsity is weak, ℓ_1 -norm regularization cannot characterize noise distribution characteristics effectively. It is noted that DSFCM_N achieves segmentation results close to those of WRFCM when the low level of impulse noise is involved. However, as the impulse noise level becomes high, DSFCM_N does not

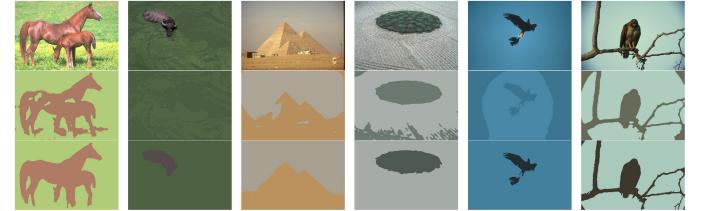


Fig. 5. Segmentation results on six real-world images. From top to bottom: observed images and results of DSFCM_N and WRFCM.

TABLE III
COMPARISON RESULTS ON REAL-WORLD IMAGES

Image	DSFCM_N			WRFCM		
	SA	SDS	MCC	SA	SDS	MCC
Fig. 5 column 1	93.116	90.279	84.987	98.732	98.162	97.201
Fig. 5 column 2	50.688	11.093	0.638	99.746	97.906	97.771
Fig. 5 column 3	92.101	90.791	83.922	99.442	99.520	98.857
Fig. 5 column 4	86.100	90.923	67.245	99.518	98.934	98.319
Fig. 5 column 5	58.749	18.582	24.097	99.639	99.003	99.856
Fig. 5 column 6	94.433	86.987	84.502	98.423	94.369	93.571

TABLE IV
COMPUTATIONAL COMPLEXITY OF DSFCM_N AND WRFCM

Algorithm	Computational complexity
DSFCM_N	$K \times c \times t \times \mathcal{N}_i $
WRFCM	$K \times c \times t \times \mathcal{N}_i $

work well since the percentage of damaged pixels is high. Superior to DSFCM_N, WRFCM takes analyzed noise distribution characteristics as a starting point, thus producing better segmentation results.

B. Comparison Results on Real-World Images

To further illustrate WRFCM's superiority over DSFCM_N, we also test real-world images. We randomly choose six examples from the public BSDS500 database.¹ They contain unknown noise, outliers, intensity inhomogeneity, etc. The visual results of DSFCM_N and WRFCM are portrayed in Fig. 5. Correspondingly, the quantitative results are listed in Table III. From Fig. 5 and Table III, we see that WRFCM outperforms DSFCM_N apparently, since it preserves more contours and feature details. In a quantitative fashion, WRFCM's average SA improvement over DSFCM is 25.32%.

C. Comparison Results on Computing Overheads

We provide a comparison of the computational overhead of DSFCM_N and WRFCM. Their computational complexity is presented in Table IV. It is noted that K is the number of image pixels, c is the number of prototypes, t is the iteration count, and $|\mathcal{N}_i|$ is the size of a local window \mathcal{N}_i .

As Table IV shows, the two algorithms have the same computational complexity, that is, $O(K)$. In addition, to compare their practicability, we present their comparison results in terms of computing time (in seconds), as summarized in

¹<https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>

TABLE V
COMPARISON RESULTS ON COMPUTING TIME (IN SECONDS)

Algorithm	Fig. 4 column 1	Fig. 4 column 2	Fig. 4 column 3	Fig. 4 column 4	Fig. 4 column 5	Fig. 5 (Average)
DSFCM_N	6.63	10.68	5.90	7.33	30.37	11.72
WRFCM	0.83	1.78	2.24	2.94	5.27	2.79

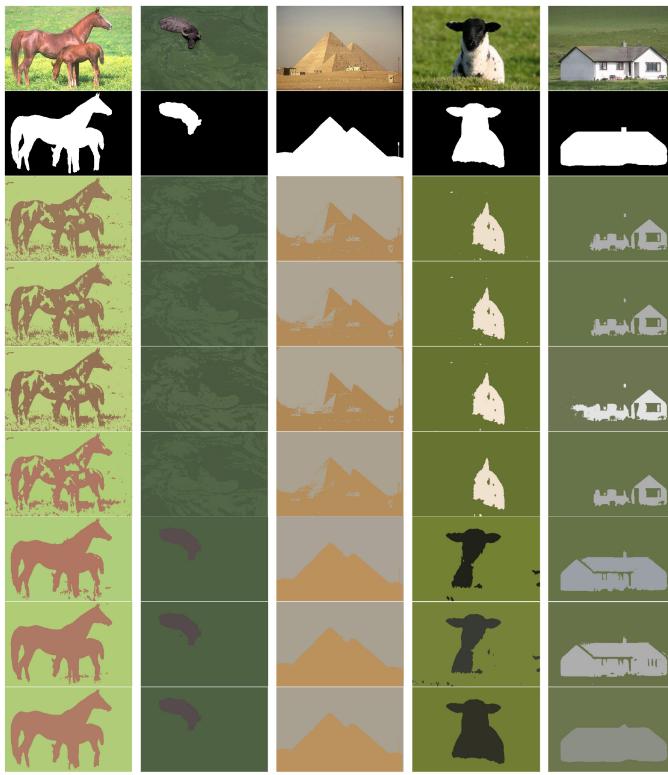


Fig. 6. Segmentation results on five real-world images in BSDS500 and MSRC. From top to bottom: observed images, ground truth, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, RFRCM, WFCM, and WRFCM.

Table V. All experiments are implemented with MATLAB running on a laptop with Intel Core i5-8250U CPU of (1.60 GHz) and 8.0-GB RAM. Table V shows that DSFCM_N is more time consuming than WRFCM since it requires more iterations. In conclusion, WRFCM runs faster than DSFCM_N despite they have the same theoretical computational complexity.

D. Comparison With Other Baseline Methods

To further verify WRFCM's superiority, we offer the comparison results with six baseline methods, that is, FCM_S1 [6], FCM_S2 [6], FLICM [9], KWFLICM [11], RFRCM [28], and WFCM [18]. Such results are produced by testing real-world images in two public databases, that is, BSDS500 and MSRC.² The average segmentation results of all algorithms are summarized in shown in Table VI. The optimal values are bolded. Sampled visual results are shown in Fig. 6.

²<http://research.microsoft.com/vision/cambridge/recognition/>

TABLE VI
AVERAGE SEGMENTATION PERFORMANCE (%) ON REAL-WORLD IMAGES IN BSDS500 AND MSRC

Algorithm	BSDS500			MSRC		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	68.223	76.101	34.887	79.703	67.928	49.122
FCM_S2	68.225	50.438	34.700	79.511	67.003	48.569
FLICM	77.000	83.664	50.134	80.694	51.158	51.879
KWFLICM	66.663	73.264	34.552	79.224	45.757	47.766
RFRCM	98.924	98.379	96.982	90.480	93.894	75.835
WFCM	99.142	98.916	97.462	90.463	78.304	75.881
WRFCM	99.307	98.529	97.943	99.848	99.839	99.384

As shown in Fig. 6 and Table VI, WRFCM is obviously superior to its peers. In a quantitative fashion, WRFCM's average SA improvements over its peers are 45.56%, 45.56%, 28.97%, 48.97%, 0.39%, and 0.17%, respectively, when segmenting 500 real-world images in the BSDS500 database. As to the MSRC database, the improvements are 25.28%, 25.58%, 23.74%, 26.03%, 10.35%, and 10.37%, respectively. FCM_S1, FCM_S2, FLICM, and KWFLICM achieve unsatisfactory segmentation results. They only consider noise suppression without analyzing the difference between noise and image features. Compared with them, RFRCM and WFCM preserve more contours and feature details. Due to the use of noise estimation, WRFCM analyzes noise characteristics well and separates the noise from image features. As a result, it widely acquires optimal SA, SDS, and MCC values much more than its peers.

VII. CONCLUSION AND FUTURE WORK

In this article, a comparative study of two NFCM methods, namely: 1) DSFCM and 2) RFRCM, was completed. Three issues were carefully investigated: 1) RFRCM realizes more accurate noise estimations DSFCM when different types of noise are encountered; 2) DSFCM is a particular case of RFRCM when only impulse noise is involved; and 3) RFRCM has higher effectiveness and execution efficiency than DSFCM. This study developed FCMs for image segmentation, since it illustrates the interpretability of the RFRCM framework well and provides more options for the inclusion of the regularization term in NFCM.

There are some open issues worth pursuing as future work. First, DSFCM's deficiencies can be overcome by using a transform method, such as Fourier transform; orthonormal wavelets [32]; translation-invariant wavelets [33]; and framelets [34], [35], [36], [37]. In some transformed domain, noise/outliers are sparse and generally located in high-frequency channels. Based on the transform, DSFCM can be applied. Second, as a competing approach for image segmentation, can WRFCM be extended to other research fields, such as remote sensing [67]; ecological systems [68]; transportation networks [69], [70]; and other applications [71], [72], [73], [74], [75]. How can the number of clusters be selected automatically? Answering them calls for more research effort.

REFERENCES

- [1] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.
- [2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
- [3] M. Yambal and H. Gupta, "Image segmentation using fuzzy C-means clustering: A survey," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 7, pp. 2927–2929, Jul. 2013.
- [4] P. K. Mishro, S. Agrawal, R. Panda, and A. Abraham, "A novel type-2 fuzzy C-means clustering for brain MR image segmentation," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3901–3912, Aug. 2021.
- [5] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.
- [6] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.
- [7] L. Szilagyi, Z. Benyo, S. Szilagyi, and H. Adam, "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in *Proc. IEEE 25th Annu. Int. Conf. EMBS*, Sep. 2003, pp. 724–726.
- [8] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.
- [9] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, pp. 1328–1337, 2010.
- [10] T. Celik and H. K. Lee, "Comments on 'A robust fuzzy local information C-means clustering algorithm,'" *IEEE Trans. Image Process.*, vol. 22, pp. 1258–1261, 2013.
- [11] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, pp. 573–584, 2013.
- [12] K. P. Lin, "A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1074–1087, Oct. 2014.
- [13] A. Elazab, C. Wang, F. Jia, J. Wu, G. Li, and Q. Hu, "Segmentation of brain tissues from magnetic resonance images using adaptively regularized kernel-based fuzzy-means clustering," *Comput. Math. Method. Med.*, vol. 2015, pp. 1–12, Dec. 2015.
- [14] F. Zhao, L. Jiao, and H. Liu, "Kernel generalized fuzzy C-means clustering with spatial information for image segmentation," *Digit. Signal Process.*, vol. 23, no. 1, pp. 184–199, 2013.
- [15] F. Guo, X. Wang, and J. Shen, "Adaptive fuzzy C-means algorithm based on local noise detecting for image segmentation," *IET Image Process.*, vol. 10, no. 4, pp. 272–279, 2016.
- [16] Z. Zhao, L. Cheng, and G. Cheng, "Neighbourhood weighted fuzzy c-means clustering algorithm for image segmentation," *IET Image Process.*, vol. 8, no. 3, pp. 150–161, 2014.
- [17] X. Zhu, W. Pedrycz, and Z. W. Li, "Fuzzy clustering with nonlinearly transformed data," *Appl. Soft Comput.*, vol. 61, pp. 364–376, Dec. 2017.
- [18] C. Wang, W. Pedrycz, J. Yang, M. Zhou, and Z. Li, "Wavelet frame-based fuzzy C-means clustering for segmenting images on graphs," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3938–3949, Sep. 2020.
- [19] R. R. Ghareeb, G. Gendy, A. Abdelfattah, and H. Selim, "Adaptive local data and membership based KL divergence incorporating C-means algorithm for fuzzy image segmentation," *Appl. Soft Comput.*, vol. 59, pp. 143–152, Oct. 2017.
- [20] C. Wang, W. Pedrycz, Z. Li, and M. Zhou, "Kullback-Leibler divergence-based Fuzzy C-Means clustering incorporating morphological reconstruction and wavelet frames for image segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7612–7623, Aug. 2022.
- [21] C. Wang, W. Pedrycz, Z. Li, M. Zhou, and S. S. Ge, "G-image segmentation: Similarity-preserving fuzzy C-means with spatial information constraint in wavelet space," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3887–3898, Dec. 2021.
- [22] J. Gu, L. Jiao, S. Yang, and F. Liu, "Fuzzy double C-means clustering based on sparse self-representation," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 612–626, Apr. 2018.
- [23] C. Wang, W. Pedrycz, M. Zhou, and Z. Li, "Sparse regularization-based Fuzzy C-Means clustering incorporating morphological grayscale reconstruction and wavelet frames," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 7, pp. 1826–1840, Jul. 2021.
- [24] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, pp. 176–201, 1993.
- [25] J. Chen, C. Su, W. Grimson, J. Liu, and D. Shiue, "Object segmentation of database images by dual multiscale morphological reconstruction and retrieval applications," *IEEE Trans. Image Process.*, vol. 21, pp. 828–843, 2012.
- [26] T. Lei, P. Liu, X. Jia, X. Zhang, H. Meng, and A. K. Nandi, "Automatic fuzzy clustering framework for image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2078–2092, Sep. 2020.
- [27] C. Wang, W. Pedrycz, Z. Li, M. Zhou, and J. Zhao, "Residual-sparse Fuzzy C-means clustering incorporating morphological reconstruction and wavelet frame," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3910–3924, Dec. 2021.
- [28] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and K. N. Asoke, "Significantly fast and robust fuzzy C-means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, Oct. 2018.
- [29] X. Bai, Y. Zhang, H. Liu, and Z. Chen, "Similarity measure-based possibilistic FCM with label information for brain MRI segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2618–2630, Jul. 2019.
- [30] Y. Zhang, X. Bai, R. Fan, and Z. Wang, "Deviation-sparse fuzzy C-means with neighbor information constraint," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 185–199, Jan. 2019.
- [31] C. Wang, W. Pedrycz, Z. Li, and M. Zhou, "Residual-driven Fuzzy C-Means clustering for image segmentation," *IEEE/CAC J. Automatica Sinica*, vol. 8, no. 4, pp. 876–889, Apr. 2021.
- [32] I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF Regional Conference Series in Applied Mathematics), vol. 61. Philadelphia, PA, USA: SIAM, 1992.
- [33] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York, NY, USA: Springer-Verlag, 1995, pp. 125–150.
- [34] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: MRA-based constructions of wavelet frames," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 1, pp. 1–46, Jan. 2003.
- [35] C. Wang and J. Yang, "Poisson noise removal of images on graphs using tight wavelet frames," *Visual Comput.*, vol. 34, no. 10, pp. 1357–1369, Oct. 2018.
- [36] J. Yang and C. Wang, "A wavelet frame approach for removal of mixed Gaussian and impulse noise on surfaces," *Inverse Probl. Imag.*, vol. 11, no. 5, pp. 783–798, Oct. 2017.
- [37] C. Wang, Z. Yan, W. Pedrycz, M. Zhou, and Z. Li, "A weighted fidelity and regularization-based method for mixed or unknown noise removal from images on graphs," *IEEE Trans. Image Process.*, vol. 29, pp. 5229–5243, 2020.
- [38] J. Jiang, L. Zhang, and J. Yang, "Mixed noise removal by weighted encoding with sparse nonlocal regularization," *IEEE Trans. Image Process.*, vol. 23, pp. 2651–2662, 2014.
- [39] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan, "Tensor low-rank representation for data recovery and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1718–1732, May 2021.
- [40] A. Fakhry, T. Zeng, and S. Ji, "Residual deconvolutional networks for brain electron microscopy image segmentation," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 447–456, Feb. 2017.
- [41] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, pp. 3142–3155, 2017.
- [42] F. Kokkinos and S. Lefkimmatis, "Iterative residual CNNs for burst photography applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5929–5938.
- [43] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M. H. Yang, "Simultaneous fidelity and regularization learning for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 284–299, Jan. 2021.
- [44] Y. Zhang, X. Li, M. Lin, B. Chiu, and M. Zhao, "Deep-recursive residual network for image semantic segmentation," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12935–12947, Jan. 2020.
- [45] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, nos. 1–2, pp. 89–97, Jan. 2004.
- [46] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
- [47] C. Clason, B. Jin, and K. Kunisch, "A duality-based splitting method for ℓ^1 -TV image restoration with automatic regularization parameter choice," *SIAM J. Sci. Comput.*, vol. 32, no. 3, pp. 1484–1505, May 2010.

- [48] J. Yang, Y. Zhang, and W. Yin, "An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise," *SIAM J. Sci. Comput.*, vol. 31, no. 4, pp. 2842–2865, Jul. 2009.
- [49] T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," *J. Math. Imag. Vis.*, vol. 27, no. 3, pp. 257–263, Apr. 2007.
- [50] G. Steidl and T. Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods," *J. Math. Imag. Vis.*, vol. 36, no. 2, pp. 168–184, 2010.
- [51] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, pp. 1413–1457, Nov. 2004.
- [52] J. Sun and Z. Xu, "Color image denoising via discriminatively learned iterative shrinkage," *IEEE Trans. Image Process.*, vol. 24, pp. 4148–4159, 2015.
- [53] V. W. Macrelo, S. H. Elias, and R. P. Daniel, "Accelerating overrelaxed and monotone fast iterative shrinkage-thresholding algorithms with line search for sparse reconstructions," *IEEE Trans. Image Process.*, vol. 26, pp. 3569–3578, 2017.
- [54] G. Corliss, "Which root does the bisection algorithm find?" *SIAM Rev.*, vol. 19, no. 2, pp. 325–327, 1977.
- [55] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 1984.
- [56] C. Clason, " L^∞ fitting for inverse problems with uniform noise," *Inverse Problems*, vol. 28, no. 10, Oct. 2012, Art. no. 104007.
- [57] P. Weiss, G. Aubert, and L. Blanc-Féraud, "Some application of L^∞ -constraints in image processing," INRIA Res., Le Chesnay Cedex, France, Rep. RR-6115, Nov. 2006.
- [58] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM J. Appl. Math.*, vol. 68, no. 4, pp. 925–946, Jan. 2008.
- [59] H. Woo and S. Yun, "Proximal linearized alternating direction method for multiplicative denoising," *SIAM J. Sci. Comput.*, vol. 35, no. 2, pp. B336–B358, Mar. 2013.
- [60] M. V. Afonso and J. M. R. Sanches, "Blind inpainting using ℓ_0 and total variation regularization," *IEEE Trans. Image Process.*, vol. 24, pp. 2239–2253, 2015.
- [61] J. Seabra, J. Xavier, and J. Sanches, "Convex ultrasound image reconstruction with log-Euclidean priors," in *Proc. Eng. Med. Biol. Soc. (EMBS)*, Aug. 2008, pp. 435–438.
- [62] M. Yan, "Restoration of images corrupted by impulse noise and mixed Gaussian impulse noise using blind inpainting," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1227–1245, Jul. 2013.
- [63] G. Yuan and B. Ghanem, " ℓ_0 TV: A sparse optimization method for impulse noise image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 352–364, Feb. 2019.
- [64] D. N. H. Thanh, D. Sergey, V. B. S. Prasath, and N. H. Hai, "Blood vessels segmentation method for retinal fundus images based on adaptive principal curvature and image derivative operators," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W12, pp. 211–218, May 2019, doi: [10.5194/isprs-archives-XLII-2-W12-211-2019](https://doi.org/10.5194/isprs-archives-XLII-2-W12-211-2019).
- [65] D. N. H. Thanh, U. Erkan, V. B. S. Prasath, V. Kumar, and N. N. Hien, "A skin lesion segmentation method for dermoscopic images based on adaptive thresholding with normalization of color models," in *Proc. IEEE 6th Int. Conf. Electr. Electron. Eng.*, Apr. 2019, pp. 116–120.
- [66] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 29, pp. 1–29, Aug. 2015.
- [67] T. Xu, L. Jiao, and W. J. Emery, "SAR image content retrieval based on fuzzy similarity and relevance feedback," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1824–1842, May 2017.
- [68] C. Wang, J. Chen, Z. Li, E. Nasr, and A. M. El-Tamimi, "An indicator system for evaluating the development of land-sea coordination systems: A case study of Lianyungang port," *Ecol. Indicators*, vol. 98, pp. 112–120, Mar. 2019.
- [69] Y. Lv, Y. Chen, X. Zhang, Y. Duan, and N. Li, "Social media based transportation research: The state of the work and the networking," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 1, pp. 19–26, Jan. 2017.
- [70] H. Han, M. Zhou, and Y. Zhang, "Can virtual samples solve small sample size problem of KISSME in pedestrian re-identification of smart transportation?" *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3766–3776, Sep. 2020.
- [71] W. Hu, Y. Huang, F. Zhang, and R. Li, "Noise-tolerant paradigm for training face recognition CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11887–11896.
- [72] Z. Cao, X. Xu, B. Hu, M. Zhou, and Q. Li, "Real-time gesture recognition based on feature recalibration network with multi-scale information," *Neurocomputing*, vol. 347, pp. 119–130, Jun. 2019.
- [73] N. Yang, B. Xia, Z. Han, and T. Wang, "A domain-guided model for facial cartoonization," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 10, pp. 1886–1888, Oct. 2022.
- [74] E. Q. Wu et al., "Self-paced dynamic infinite mixture model for fatigue evaluation of pilots' brains," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 5623–5638, Jul. 2022.
- [75] G. Cai, L. He, M. Zhou, H. Alhumade, and D. Hu, "Learning smooth representation for unsupervised domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Nov. 2021, doi: [10.1109/TNNLS.2021.3119889](https://doi.org/10.1109/TNNLS.2021.3119889).



Cong Wang (Member, IEEE) received the B.S. degree in automation and the M.S. degree in mathematics from Hohai University, Nanjing, China, in 2014 and 2017, respectively, and the Ph.D. degree in mechatronic engineering from Xidian University, Xi'an, China, in 2021.

He is currently an Assistant Professor with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an. He was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, and the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. He was also a Research Assistant with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has over 30 papers and host/participated in more than 10 research projects (over 2.5 million RMB). He also has a book and a chapter to be published. He is funded by the China National Postdoctoral Program for Innovative Talents and the Excellent Chinese and Foreign Youth Exchange Program of the China Association for Science and Technology. His current research interests include information computation, G-image processing, fuzzy theory, as well as wavelet analysis and its applications.

Dr. Wang serves as a Frequent Reviewer for more than 40 international journals, including a number of the IEEE TRANSACTIONS and many international conferences. He is a member of the Institute of Electrical and Electronics Engineers, Chinese Institute of Electronics, China Computer Federation, China Society of Image and Graphics, and China Society for Industrial and Applied Mathematics.



MengChu Zhou (Fellow, IEEE) received the B.S. degree in control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1983, the M.S. degree in automatic control from the Beijing Institute of Technology, Beijing, China, in 1986, and the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined the New Jersey Institute of Technology, Newark, NJ, USA, in 1990, and has been a Distinguished Professor of Electrical and Computer Engineering since 2013. He is also affiliated with the School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou, China. He has led or participated in over 50 research and education projects with total budget over \$12M, funded by National Science Foundation, Department of Defense, NIST, New Jersey Science and Technology Commission, and industry. His recently coauthored/edited books include the *Supervisory Control and Scheduling of Resource Allocation Systems: Reachability Graph Perspective* (IEEE Press/Wiley: Hoboken, NJ, USA, 2020) (with B. Huang) and *Sustainable Manufacturing Systems: An Energy Perspective* (IEEE Press/Wiley, Hoboken, NJ, USA, 2022) (with L. Li). He has over 1000 publications, including 14 books, more than 600 journal papers (more than 500 in IEEE TRANSACTIONS), 31 patents, and 30 book chapters. His research interests are in Petri nets, intelligent automation, Internet of Things, big data analytics, artificial intelligence, Web services, and intelligent transportation.

Dr. Zhou is a recipient of the Humboldt Research Award for U.S. Senior Scientists from Alexander von Humboldt Foundation, the Franklin V. Taylor Memorial Award, and the Norbert Wiener Award from the IEEE Systems, Man and Cybernetics Society. He has been among the Most Highly Cited Scholars for years and ranked top one in the field of engineering worldwide in 2012 by the Web of Science. He is the Founding Editor of IEEE Press Book Series on Systems Science and Engineering and the Editor-in-Chief of IEEE/CAA JOURNAL OF AUTOMATICA SINICA. He served as an Associate Editor for IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and an Editor of IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He served as a Guest Editor for many journals, including IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. He is currently an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and *Frontiers of Information Technology & Electronic Engineering*. He is the Founding Chair/Co-Chair of Technical Committee on AI-Based Smart Manufacturing Systems of IEEE Systems, Man, and Cybernetics Society, Technical Committee on Semiconductor Manufacturing Automation and Technical Committee on Digital Manufacturing and Human-Centered Automation of IEEE Robotics and Automation Society. He was the General Chair of IEEE Conference on Automation Science and Engineering, Washington, DC, USA, 23–26 August 2008; the General Co-Chair of the 2003 IEEE International Conference on System, Man and Cybernetics (SMC), Washington, DC, USA, 5–8 October 2003 and the 2019 IEEE International Conference on SMC, Bari, Italy, 6–9 October 2019; the Founding General Co-Chair of the 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, 21–23 March 2004, and the General Chair of the 2006 IEEE International Conference on Networking, Sensing and Control, Ft. Lauderdale, FL, USA, 23–25 April 2006. He was the Program Chair of 2010 IEEE International Conference on Mechatronics and Automation, 4–7 August 2010, Xi'an, China, 1998, and the 2001 IEEE International Conference on SMC and the 1997 IEEE International Conference on Emerging Technologies and Factory Automation. He is a Life Member of the Chinese Association for Science and Technology—USA and served as its President in 1999. He is a Fellow of the International Federation of Automatic Control, American Association for the Advancement of Science, Chinese Association of Automation, and National Academy of Inventors.



Witold Pedrycz (Life Fellow, IEEE) received the M.Sc. degree in computer science and technology, the Ph.D. degree in computer engineering, and the D.Sc. degree in systems science from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively.

He is a Professor and the Canada Research Chair of Computational Intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. He has authored 15 research monographs covering various aspects of computational intelligence, data mining, and software engineering. His current research interests include computational intelligence, fuzzy modeling, and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in the above areas.

Dr. Pedrycz was a recipient of the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is intensively involved in editorial activities. He is an Editor-in-Chief of the *Information Sciences*, *WIREs Data Mining and Knowledge Discovery* (Wiley), and *International Journal of Granular Computing* (Springer). He currently serves as a member for a number of editorial boards of other international journals. He is a Foreign Member of the Polish academy of Sciences. He is a Fellow of the Royal Society of Canada.



ZhiWu Li (Fellow, IEEE) received the B.S. degree in mechanical engineering, the M.S. degree in automatic control, and the Ph.D. degree in manufacturing engineering from Xidian University, Xi'an, China, in 1989, 1992, and 1995, respectively.

He joined Xidian University in 1992. He is also currently with the Institute of Systems Engineering, Macau University of Science and Technology, Macau, China. He was a Visiting Professor with the University of Toronto, Toronto, ON, Canada; Technion-Israel Institute of Technology, Haifa, Israel; Martin-Luther University of Halle-Wittenburg, Halle, Germany; Conservatoire National des Arts et Métiers, Paris, France; and Meliksah Universitesi, Kayseri, Turkey. His current research interests include Petri net theory and application, supervisory control of discrete-event systems, workflow modeling and analysis, system reconfiguration, game theory, and data and process mining.

Dr. Li was a recipient of the Alexander von Humboldt Research Grant, Alexander von Humboldt Foundation, Germany. He is listed in Marquis Who's Who in the World, 27th Edition, 2010. He serves as a Frequent Reviewer for more than 90 international journals, including *Automatica* and a number of IEEE TRANSACTIONS, as well as many international conferences. He is the Founding Chair of the Xi'an Chapter of IEEE Systems, Man, and Cybernetics Society. He is a member of the Discrete-Event Systems Technical Committee of the IEEE Systems, Man, and Cybernetics Society and IFAC Technical Committee on Discrete-Event and Hybrid Systems from 2011 to 2014.